PCT

(51) International Patent Classification 7 : H04N 7/26	A1	(11) International Publication Number: WO 00/51357 (43) International Publication Date: 31 August 2000 (31.08.00)

(21) International Application Number: PCT/US00/04824 (22) International Filing Date: 25 February 2000 (25.02.00) (30) Priority Data: 60/121,531 25 February 1999 (25.02.99) US 09/335,810 18 June 1999 (18.06.99) US (71) Applicant: SARNOFF CORPORATION [US/US]; 201 Washington Road, CN5300, Princeton, NJ 08543-5300 (US). · (72) Inventor: LEE, Jungwoo; 4051 Bayberry Court, Monmouth Junction, NJ 08852 (US). (74) Agents: NEY, Andrew, L.; Ratner & Prestia, 301 One Westlakes (Berwyn), P.O. Box 980, Valley Forge, PA 19482-0980 (US) et al.	(81) Designated States: BR, CN, JP, KR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published *With international search report.*

(54) Title: TRANSCODING BETWEEN DIFFERENT DCT-BASED IMAGE COMPRESSION STANDARDS
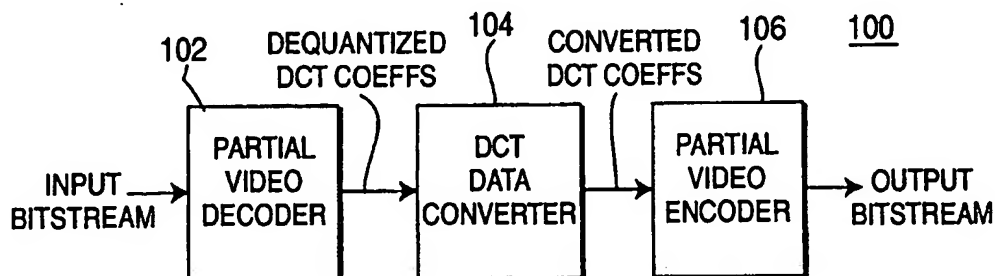
(57) Abstract

An input encoded video bitstream conforming to a first DCT-based image compression standard (e.g., MJPEG or DV) is converted into an output encoded video bitstream conforming to a second DCT-based image compression standard (e.g., MPEG) without having to fully decode the first bitstream according to the first image compression standard and without having to fully encode according to the second image compression standard in order to generate the output bitstream. Rather, a partial video decoder of a transcoding system applies steps of the first image compression standard to the input bitstream to generate dequantized DCT coefficient data. A DCT data converter of the transcoding system then processes the dequantized DCT coefficient data in the DCT domain as necessary to enable a partial video encoder of the transcoding system to apply steps of the second image compression standard to the converted DCT coefficient data in order to generate the output bitstream.

- 1 -

TRANSCODING BETWEEN
DIFFERENT DCT-BASED IMAGE COMPRESSION STANDARDS
BACKGROUND OF THE INVENTION

Field of the Invention

5   The present invention relates to image processing, and, in particular, to transcoding encoded image data between bitstreams conforming to two different image compression standards.

10

Description of the Related Art

A number of different image compression standards have been and will continue to be used to encode image data for more efficient storage and/or transmission of video content. The JPEG (Joint Photographic Experts Group) standard was originally designed for still images, but is also applied to sequences of images in which each image is encoded using only intra-frame encoding techniques (i.e., without reference to any other images in the sequence). Such encoded data is referred to as motion-JPEG or MJPEG encoded data. In MJPEG encoding, each image is transformed using a block-based discrete cosine transform (DCT). The resulting DCT coefficients are then quantized and run-length encoded to generate sets of run-length pairs. The run-length pairs are then encoded into an MJPEG bitstream using variable-length coding (VLC).

The DV (Digital Video) standard is a coding standard for digital video camcorders and digital VCRs. Like MJPEG, the DV standard relies primarily on DCT-based intra-frame encoding techniques to encode sequences of images. One major difference between the DV and MJPEG standards is that DV encoding supports two different modes for the DCT transform: a frame mode and a field mode. In the frame mode, also referred to as the 8x8 mode, 8x8 blocks of pixel data are encoded using an 8x8 DCT transform, similar to the processing in MJPEG encoding.

- 2 -

In the field mode, also referred to as the 2-4x8 mode, image data are encoded using a 4x4 DCT transform. In 2-4x8 mode, two different types of 4x8 blocks of DCT coefficient data are generated: even 4x8 DCT blocks and odd 4x8 DCT blocks. An even 4x8 DCT block X_e corresponds to the DCT of the sum of pixel data from the top and bottom fields of a video frame, as represented in the following Equation (1):

$$X_e = C_4 [x(2i,j) + x(2i+1,j)] C_8', \tag{1}$$

where C_4 is a length-4 DCT transform matrix, C_8' is the transpose of a length-8 DCT transform matrix, $x(2i,j)$ is a 4x8 block of pixel data from the top field, and $x(2i+1,j)$ is the corresponding 4x8 block of pixel data from the bottom field. Similarly, the corresponding odd 4x8 DCT block X_o corresponds to the DCT of the difference of the pixel data from the same top and bottom fields, as represented in the following Equation (2):

$$X_o = C_4 [x(2i,j) - x(2i+1,j)] C_8'. \tag{2}$$

The even and odd sets of DCT coefficient data generated using Equations (1) and (2) are then quantized, run-length encoded, and variable-length encoded.

The MPEG (Moving Picture Experts Group) standard was designed for sequences of images, in which each image is encoded using intra-frame encoding techniques and/or inter-frame encoding techniques (in which image data are encoded based on pixel differences between the current image and a reference image that is generated from one or more other images in the sequence). As in MJPEG and DV processing, in MPEG processing, a DCT transform is applied to blocks of image data to generate blocks of DCT coefficients that are then further processed (i.e., quantized, run-length encoded, and variable-length encoded) to generate the corresponding MPEG encoded bitstream.

Much encoded video content exists -- and will continue to be generated -- based on the MJPEG and DV standards. It would be advantageous to be able to make such MJPEG- and DV-based video content available to, for example, PC users having only MPEG image processors. This would enable someone with a DV-based camcorder and an MPEG-based PC to generate video content with the camcorder and then play and otherwise process that video content on the PC.

- 3 -

Transcoding refers to the process of converting an input encoded video bitstream that conforms to one image processing standard (e.g., MJPEG or DV) into an output encoded video bitstream that conforms to another image processing standard (e.g., MPEG). One brute-force approach to transcoding is to fully decode the input bitstream using a decoder

5   conforming to the first image processing standard and then re-encode the resulting decoded sequence of images using an encoder conforming to the second image processing standard. In order to implement such brute-force transcoding in many real-time applications (i.e., where the transcoder is required to generate the output bitstream at the same frame rate at which it receives the input bitstream), the transcoders would need to be implemented using

10   expensive hardware-based (e.g., MJPEG or DV) decoders and (e.g., MPEG) encoders.

SUMMARY OF THE INVENTION

The present invention is directed to techniques for transcoding input encoded video bitstreams conforming to a first DCT-based image compression standard (e.g., MJPEG or DV) into output encoded video bitstreams conforming to a second DCT-based image

15   compression standard (e.g., MPEG). As opposed to brute-force transcoding techniques in which an input bitstream is fully decoded and then fully re-encoded to generate an output bitstream, under the present invention, the input bitstream is only partially decoded according to the first standard into the DCT domain (i.e., dequantized DCT coefficients), and then the re-encoding processing for the second standard starts with those dequantized

20   DCT coefficients to generate the output bitstream. Because transcoders of the present invention only perform part of the full decoding and encoding processes, which do not include application of the computationally expensive inverse and forward DCT transforms, these transcoders can be implemented using PC-based software-only solutions and still meet the throughput requirements of many real-time applications. As such, the expense of

25   requiring full decoders and encoders is avoided by the present invention.

According to one embodiment, the present invention is a method for converting an input encoded video bitstream conforming to a first DCT-based compression algorithm into an output encoded video bitstream conforming to a second DCT-based compression algorithm different from the first compression algorithm, comprising the steps of (a)

30   applying decoding steps conforming to the first compression algorithm to the input

- 4 -

bitstream to generate dequantized DCT coefficient data in a DCT domain; (b) subsampling and anti-aliasing filtering the dequantized DCT coefficient data in the DCT domain in order to resize the dequantized DCT coefficient data; and (c) applying encoding steps conforming to the second compression algorithm to the resized DCT coefficient data to generate the output bitstream.

According to another embodiment, the present invention is a method for converting an input encoded video bitstream conforming to a first DCT-based compression algorithm into an output encoded video bitstream conforming to a second DCT-based compression algorithm different from the first DCT-based compression algorithm, comprising the steps of (a) applying decoding steps conforming to the first compression algorithm to the input bitstream to generate dequantized DCT coefficient data in a DCT domain; (b) converting the dequantized DCT coefficient data from a 2-4x8 mode format into an 8x8 mode format in the DCT domain; and (c) applying encoding steps conforming to the second compression algorithm to the converted DCT coefficient data in the 8x8 mode format to generate the output bitstream.

According to yet another embodiment, the present invention is a method for converting an input encoded video bitstream conforming to a first DCT-based compression algorithm into an output encoded video bitstream conforming to a second DCT-based compression algorithm different from the first DCT-based compression algorithm, comprising the steps of (a) applying decoding steps conforming to the first compression algorithm to the input bitstream to generate dequantized DCT coefficient data in a DCT domain; (b) converting the dequantized DCT coefficient data from a YUV 4:1:1 color format into a YUV 4:2:0 color format in the DCT domain; and (c) applying encoding steps conforming to the second compression algorithm to the converted DCT coefficient data in the YUV 4:2:0 color format to generate the output bitstream.

According to still another embodiment, the present invention is a method for converting an input encoded video bitstream conforming to a first DCT-based compression algorithm into an output encoded video bitstream conforming to a second DCT-based compression algorithm different from the first DCT-based compression algorithm, comprising the steps of (a) applying decoding steps conforming to the first compression

algorithm to the input bitstream to generate dequantized DCT coefficient data in a DCT domain; (b) performing motion-compensated inter-frame differencing on the dequantized DCT coefficient data in the DCT domain based on motion vectors corresponding to block boundaries; and (c) applying encoding steps conforming to the second compression

5   algorithm to the motion-compensated inter-frame DCT coefficient difference data to generate the output bitstream.

According to still yet another embodiment, the present invention is a method for converting an input encoded video bitstream conforming to a first DCT-based compression algorithm into an output encoded video bitstream conforming to a second DCT-based

10   compression algorithm different from the first DCT-based compression algorithm, comprising the steps of (a) applying decoding steps conforming to the first compression algorithm to the input bitstream to generate dequantized DCT coefficient data in a DCT domain; (b) performing a field/frame mode decision on the dequantized DCT coefficient data in the DCT domain to determine whether to encode the dequantized DCT coefficient

15   data in a field mode or in a frame mode; and (c) applying encoding steps conforming to the second compression algorithm to the DCT coefficient data in one of the field mode and the frame mode to generate the output bitstream.

BRIEF DESCRIPTION OF THE DRAWINGS

Other aspects, features, and advantages of the present invention will become more

20   fully apparent from the following detailed description, the appended claims, and the accompanying drawings in which:

Fig. 1 shows a block diagram of a transcoding system, according to one embodiment of the present invention;

Figs. 2A-C and 3A-C show the relative sizes for corresponding regions of pixel data

25   for the YUV 4:1:1 and YUV 4:2:0 color formats; and

Fig. 4 shows a flow diagram of the DV-to-MPEG transcoding processing performed by transcoding system of Fig. 1, according to one embodiment of the present invention.

- 6 -

DETAILED DESCRIPTION

There are three different aspects to transcoding: syntax conversion, image resizing, and bit-rate control. Syntax conversion refers to the rearranging of data in the input encoded video bitstream to ensure that the output encoded video bitstream conforms to the syntax of the corresponding image compression standard. Image resizing refers to the scaling up or down of the image size of the data in the input bitstream for the output bitstream. Image resizing may be implemented to achieve desired or perhaps required dimensions of images for the output bitstream. In addition, image resizing may be implemented in order to conform to bit-rate limitations in the output bitstream. For example, the DV standard can support bit rates as high as 25 Mbits per second, while the MPEG-2 standard is limited to 15 Mbits per second. In addition to using image resizing to affect bit rate, further processing may be required for bit-rate control. For example, if the output bitstream has tighter storage and/or transmission constraints than the input bitstream, the transcoding process may involve quantization at a greater level in the output bitstream than that used in the input bitstream in order to control the bit rate at a level well below the upper limit supported by image processing standard of the output bitstream (e.g., even less than 15 Mbits per second for MPEG-2).

Fig. 1 shows a block diagram of a transcoding system 100, according to one embodiment of the present invention. Transcoding system 100 converts an input encoded video bitstream conforming to a first image compression standard (e.g., MJPEG or DV) into an output encoded video bitstream conforming to a second image compression standard (e.g., MPEG). Transcoding system 100 comprises a partial video decoder 102, a DCT data converter 104, and a partial video encoder 106. Depending on the particular implementation, the various components of transcoding system 100 can be implemented in software, hardware, or a combination of both.

In particular, partial video decoder 102 applies processing conforming to the first image processing standard (e.g., variable-length decoding, run-length decoding, and dequantization) to partially decode the input encoded video bitstream into dequantized DCT coefficient data. DCT data converter 104 converts the dequantized DCT coefficient data generated by partial video decoder 102 into corresponding converted DCT coefficient data

- 7 -

for partial video encoder 106, which applies processing conforming to the second image compression standard (e.g., quantization, run-length encoding, and variable-length encoding) to the converted DCT coefficient data to generate the output encoded video bitstream.

5   The processing performed by partial video decoder 102 may comprise any suitable conventional processing techniques conforming to the first image processing standard in order to generate the decoded DCT coefficient data from the input encoded video bitstream. The processing performed by DCT data converter 104 in accordance with the present invention will depend on the particular video standards of the input and output encoded

10   video bitstreams. Depending on the implementation, some or all of the processing of DCT data converter 104 may be optional. Depending on the implementation, the processing performed by partial video encoder 106 may comprise any suitable conventional processing techniques conforming to the second image processing standard or it may comprise special processing techniques in accordance with the present invention. The processing of DCT

15   data converter 104 and those portions of the processing of partial video encoder 106 that are part of the present invention are described in the following sections.

DCT Data Conversion for MJPEG-to-MPEG Transcoding

           This section describes the processing of DCT data converter 104 of transcoder 100 of Fig. 1, when the input encoded video bitstream is an MJPEG bitstream and the output

20   encoded video bitstream is an MPEG bitstream, according to an embodiment of the present invention where the image size is reduced by a factor of 2 in both the horizontal and vertical dimensions. It will be understood that the image resizing of the present invention can be implemented for image resizing factors other than 2, including scaling up instead of scaling down, as well as scaling by two different factors in the horizontal and vertical dimensions.

25   The image resizing processing performed by DCT data converter 104 for MJPEG-to-MPEG transcoding can be represented by the matrix multiplication formula of Equation (3) as follows:

$$Y = TXT^t, \tag{3}$$

where X is a 16x16 block of dequantized DCT coefficients formed by concatenating a 2x2

30   region of four 8x8 blocks of dequantized DCT coefficients generated by partial video

- 8 -

decoder 102, T is an 8x16 transformation matrix, T^t is the 16x8 transpose of T, and Y is the resulting 8x8 block of converted DCT coefficients.

The transformation matrix T can be represented by the matrix multiplication formula of Equation (4) as follows:

$$T = C_8 SA \tag{4}$$

where C_8 is an 8x8 DCT transform, S is an 8x16 subsampling and anti-aliasing filter matrix, and the 16x16 matrix A can be represented by Equation (5) as follows:

$$A = \begin{bmatrix} C_8' & O_8 \\ O_8 & C_8' \end{bmatrix} \tag{5}$$

where C_8' is the transpose of C_8 and O_8 is an 8x8 zero submatrix.

The 8x16 subsampling and anti-aliasing filter matrix S can be presented by Equation (6) as follows:

$$S = \begin{bmatrix} w(0) \\ w(2) \\ \cdot \\ \cdot \\ \cdot \\ w(12) \\ w(14) \end{bmatrix} \tag{6}$$

where each w(i) is a 1x16 row of matrix S consisting of the anti-aliasing low-pass-filter kernel w shifted right by i pixels. Typical kernels are [1 0], [½ ½], and [¼ ½ ¼]. For example, when the kernel w is [½ ½], the 8 rows w(0) through w(14) of the matrix S are given be Equations (7)-(14) as follows:

$$w(0) = [½ ½ 0 0 0 0 0 0 0 0 0 0 0 0 0 0] \tag{7}$$

$$w(2) = [0 0 ½ ½ 0 0 0 0 0 0 0 0 0 0 0 0] \tag{8}$$

$$w(4) = [0 0 0 0 ½ ½ 0 0 0 0 0 0 0 0 0 0] \tag{9}$$

$$w(6) = [0 0 0 0 0 0 ½ ½ 0 0 0 0 0 0 0 0] \tag{10}$$

$$w(8) = [0 0 0 0 0 0 0 0 ½ ½ 0 0 0 0 0 0] \tag{11}$$

$$w(10) = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \tfrac{1}{2}\ \tfrac{1}{2}\ 0\ 0\ 0\ 0] \qquad (12)$$

$$w(12) = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \tfrac{1}{2}\ \tfrac{1}{2}\ 0\ 0] \qquad (13)$$

$$w(14) = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \tfrac{1}{2}\ \tfrac{1}{2}] \qquad (14)$$

Since T and T^t are independent of the DCT coefficient data, they can be pre-computed to save computations during real-time processing.

Conversion of DCT coefficients by performing the matrix-based transformation of Equation (3) simultaneously achieves a number of different functions. In addition to resizing the DCT coefficient data (i.e., in this case, reducing both the horizontal and vertical dimensions by a factor of 2), the transformation of Equation (3) also simultaneously applies a spatial-domain anti-aliasing filter to the DCT coefficient data. This spatial-domain anti-aliasing filter reduces aliasing distortion that might otherwise result from downsampling, while providing computational savings by combining the subsampling processing and filtering processing into a single step.

The processing described in this section can also be applied in the context of DV-to-MPEG transcoding for encoded video data that was generated using the DV 8x8 mode.

DCT Data Conversion for DV-to-MPEG Transcoding

The following two sections describe the processing of DCT data converter 104 of transcoder 100 of Fig. 1, when the input encoded video bitstream is a DV bitstream and the output encoded video bitstream is an MPEG bitstream, according to one embodiment of the present invention.

4:1:1-to-4:2:0 Color Format Conversion for DV-to-MPEG Transcoding

One of the major differences between DV bitstreams and MPEG bitstreams is that the DV standard uses the YUV 4:1:1 color format, while the MPEG standard uses the YUV 4:2:0 color format.

Figs. 2A-C and 3A-C show the relative sizes for corresponding regions of pixel data for the YUV 4:1:1 and YUV 4:2:0 color formats. As shown in Figs. 2A-C, for an HxW luma (Y) image, there are two HxW/4 chroma (U and V) images. As shown in Figs. 3A-C, for an HxW luma image, there are two H/2xW/2 chroma images. In order for MPEG-

compliant partial video encoder 106 of Fig. 1 to be able to generate an output bitstream conforming to the MPEG standard (which does not support the 4:1:1 color format), DCT data converter 104 is capable of converting dequantized DCT coefficient data generated by DV-compliant partial video decoder 102 from the YUV 4:1:1 color format into the YUV 4:2:0 color format, which can be handled by MPEG-compliant partial video encoder 106.

In particular, the DCT coefficients for the chroma data are converted according the matrix multiplication transformation of Equation (15) as follows:

$$Y = T_c X T_r, \tag{15}$$

where X is a 16x8 block of chroma (U or V) DCT coefficient data formed by vertically concatenating two 8x8 blocks of chroma DCT coefficient data generated by partial video decoder 102 of Fig. 1, T_c is a 8x16 column-wise transformation matrix, T_r is an 8x16 row-wise transformation matrix, and Y is the resulting 8x16 block of transformed chroma DCT coefficient data.

The column-wise transformation matrix T_c is the same as transformation matrix T as represented by Equation (4). The row-wise transformation matrix T_r can be represented by the matrix multiplication formula of Equation (16) as follows:

$$T_r = C_8 U A, \tag{16}$$

where C_8 is the 8x8 DCT transform, A is the same as the 16x16 matrix as represented by Equation (5), and the 8x16 matrix U can be represented by Equation (17) as follows:

$$U = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \tag{17}$$

Since T_c and T_r are independent of the chroma DCT coefficient data, they can be pre-computed to save computations during real-time processing.

- 11 -

Conversion of chroma DCT coefficients by performing the matrix-based transformation of Equation (15) simultaneously achieves a number of functions. The transformation of Equation (15) halves the vertical dimension and doubles the horizontal dimension, thereby effectively converting each 16x8 block of chroma DCT coefficient data

5   into an 8x16 block of chroma DCT coefficient data. In addition, the use of the filter matrix S in the column-wise transformation matrix T_c provides one-dimensional spatial-domain anti-aliasing while subsampling the chroma DCT coefficient data in the vertical direction. As in the case of the resizing/filtering processing of Equation (3), this spatial-domain anti-aliasing filter reduces aliasing distortion that might otherwise result from downsampling,

10   while providing computational savings by combining the subsampling processing and filtering processing into a single step.

The chroma conversion of chroma DCT coefficient data of Equation (15) can be applied in conjunction with the resizing/filtering processing of Equation (3). In that case, Equation (3) is applied to all of the DCT coefficient data (i.e., Y, U, and V data), while

15   Equation (15 is applied to only the chroma DCT coefficient data.

2-4x8-to-8x8 Mode Conversion for DV-to-MPEG Transcoding

Another major difference between DV bitstreams and MPEG bitstreams is the possible existence of encoded video data in a DV bitstream that was encoded based on the DV 2-4x8 mode. In order for MPEG-compliant partial video encoder 106 of Fig. 1 to be

20   able to generate an output bitstream conforming to the MPEG standard (which does not support the 2-4x8 mode), DCT data converter 104 is capable of converting dequantized DCT coefficient data generated by DV-compliant partial video decoder 102 from 2-4x8 mode to 8x8 mode, which can be handled by MPEG-compliant partial video encoder 106.

In particular, DCT data converter 104 the DCT coefficients for the chroma data are

25   converted according the matrix multiplication transformation of Equation (18) as follows:

$$Y = T_1X_e + T_2X_o, \tag{18}$$

where X_e and X_o are corresponding 4x8 blocks of even and odd DCT coefficients as defined by Equations (1) and (2), respectively, and generated by partial video decoder 102 of Fig. 1. The 8x4 transformation matrices T_1 and T_2 can be represented by Equations (19) and (20)

30   as follows:

- 12 -

$$T_1 = \tfrac{1}{2} C_8 (P_e + P_o) C_4^t \qquad (19)$$

and

$$Y = \tfrac{1}{2} C_8 (P_e - P_o) C_4^t, \qquad (20)$$

where C_8 is an 8x8 DCT transform, C_4^t is the transpose of a 4x4 DCT transform, P_e is an 8x4 top-field interleaving matrix that can be defined by Equation (21) as follows:

$$P_e = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad (21)$$

and P_o is an 8x4 bottom-field interleaving matrix that can be defined by Equation (22) as follows:

$$P_o = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (22)$$

When implementing the matrix multiplication transformation of Equation (18), DCT data converter 104 of Fig. 1 converts each corresponding pair of even and odd 4x8 DCT blocks (i.e., X_e and X_o) generated by partial video decoder 102 into a single 8x8 block of DCT coefficients (Y) for processing by partial video encoder 106.

Since T_1 and T_2 are independent of the DCT coefficient data, they can be pre-computed to save computations during real-time processing.

- 13 -

Re-Encoding of DCT Coefficient Data for MJPEG/DV-to-MPEG Transcoding

The following sections describe some of processing techniques performed by partial video encoder 106 of transcoder 100 of Fig. 1, when the input encoded video bitstream is either an MJPEG or DV bitstream and the output encoded video bitstream is an MPEG

5   bitstream, according to one embodiment of the present invention. These processing techniques relate to requantization for rate-control, inter-frame encoding in the DCT domain, motion compensation in the DCT domain, inter/intra mode decision in the DCT domain, and field/frame mode decision in the DCT domain. As described below, all of these techniques involve manipulating the DCT coefficient data generated by partial video

10   decoder 102 and/or DCT data converter 104 of Fig. 1.

Requantization for Rate Control

Depending on the particular application, there may be a need to reduce the number of bits used to represent the video content in the output encoded video bitstream, even after the resizing operations of DCT data converter 104 are implemented. One way to control the bit

15   rate of the output bitstream is to adjust the quantization levels used during the requantization processing performed by partial video encoder 106. According to the MJPEG and DV standards, every block (or macroblock) in an MJPEG or DV file is encoded using the same quantization level. On the other hand, the MPEG standards allow the quantization level to vary from frame to frame and even from macroblock to macroblock

20   within a frame. Thus, during MJPEG-to-MPEG transcoding, the possibility exists for partial video encoder 106 to vary quantization level on a macroblock basis from the MJPEG quantization level used by partial video decoder 102 in generating the dequantized DCT coefficients.

According to one embodiment of the present invention, the quantization levels used

25   by partial video encoder 106 are determined using a non-linear profile rate-control algorithm. In particular, a requantized DCT coefficient y' may be represented by Equation (23) as follows:

$$y' = y * Qjpeg / (Qmpeg * 2) \tag{23}$$

- 14 -

where y is the corresponding dequantized DCT coefficient generated by partial video decoder 102 and optionally converted by DCT data converter 104, Qjpeg is the quantization level at which the original MJPEG DCT coefficient data was quantized, and Qmpeg is the selected MPEG quantization level for the current macroblock, where Qmpeg may be represented by Equation (24) as follows:

$$Qmpeg = 31 * d / r \tag{24}$$

where Qmpeg is rounded to the nearest integer (or alternatively truncated) and clipped to the range [1, 31], r is a reaction parameter that is typically set to twice the bit count of the original MJPEG (or DV) frame, and the discrepancy d in Equation (24) may be represented by Equation (25) as follows:

$$d = S - \alpha C \tag{25}$$

where C is the cumulative bit count for the original MJPEG input bitstream and S is cumulative bit count for the resulting MPEG output bitstream, where both C and S are updated for each macroblock up to the current macroblock, and α is scale factor that can be set to a value other than 1 to either increase or decrease the target bit rate.

When too many bits are being used in the output bitstream, the value of d generated by Equation (25) will become large and positive, which will cause the value of Qmpeg generated by Equation (24) to become larger, which will cause the dequantized DCT coefficient y to be quantized at a lower level in Equation (23), thereby tending to reduce the rate at which bits will be added to the output bitstream. Analogously, when too few bits are being used in the output bitstream, d will become large and negative, causing Qmpeg to become smaller, causing y to be quantized at a higher level, thereby tending to increase the rate at which bits are added to the output bitstream.

Inter-Frame Encoding in the DCT Domain

As described in the Background section, the MPEG standard supports inter-frame encoding techniques in which image data are encoded based on pixel differences between the current image and a reference image that is generated from one or more other images in the sequence. Other than the DV 2-4x8 mode, neither the DV nor the MJPEG standard supports such inter-frame encoding. Conventional inter-frame encoding techniques involve

- 15 -

the generation of inter-frame pixel differences followed by the application of a DCT transform to the inter-frame pixel difference data. However, since DCT transforms are linear operations, inter-frame differencing can be performed on the DCT coefficients after the application of the DCT transform.

5   According to certain embodiments of the present invention, partial video encoder 106 is capable of encoding the DCT coefficient data using inter-frame encoding techniques in which the inter-frame differencing is applied in the DCT domain. Depending on the particular implementation, the inter-frame differencing may be applied to the dequantized DCT coefficient data before requantization or to the requantized DCT coefficient data after 10   requantization.

           According to one embodiment, partial video encoder 106 is capable of performing inter-frame encoding in which all of the motion vectors are zero motion vectors. In this way, partial video encoder 106 is capable of generating an output encoded video bitstream having MPEG P and B frames (i.e., frames encoded using inter-frame encoding techniques) 15   as well as I frames (i.e., frames encoded using only intra-frame encoding techniques).

Motion Compensation in the DCT Domain

           In conventional inter-frame encoding techniques, motion compensation is performed on a pixel (or even sub-pixel) basis. Although such pixel-based motion compensation is possible in the DCT domain, it is computationally very expensive. In certain embodiments 20   of the present invention, partial video encoder 106 is capable to performing non-zero motion compensation, where the motion vectors are limited to corresponding to block boundaries. In this so-called block-based motion compensation, the linearity of the relationship between inter-frame pixel differencing and the DCT transform is maintained and inter-frame differencing can be implemented in the DCT domain without expensive computations and 25   without any loss of data. This block-based motion compensation will provide better encoding for those situations in which inter-frame encoding using a non-zero, block-based motion vector results in fewer bits in the output bitstream than either inter-frame encoding using a zero motion vector or intra-frame encoding for the same image data.

- 16 -

Inter/Intra Mode Decision in the DCT Domain

The MPEG standard allows individual macroblocks in P and B frames to be encoding using either inter-frame encoding or intra-frame encoding. In conventional MPEG algorithms, the inter/intra mode decision is made in the pixel domain. According to embodiments of the present invention, partial video encoder 106 makes the inter/intra mode decision in the DCT domain based on a macroblock-level variance measure Var, which may be represented by Equations (26) and (27) as follows:

$$\mathrm{Var}_k = \frac{1}{128}\left[\sum_{i=0}^{63} c_i^2\right] - \left(\frac{1}{8}c_0\right)^2 \qquad k = 1, 2, 3, 4 \tag{26}$$

$$\mathrm{Var} = \sum_{k=1}^{4} \mathrm{Var}_k \tag{27}$$

where Var_k is the corresponding block-level variance measure, c_i is i^{th} DCT coefficient in an 8x8 block of DCT coefficients numbered in conventional zig-zag fashion from $i=0$ to $i=63$. For an "intra" variance, the DCT coefficients in Equation (26) are generated from the raw pixel data, while, for an "inter" variance, the DCT coefficients are generated from inter-frame pixel differences. Equation (26) is based on Parseval's theorem, which states that the sum of squares in the pixel domain is half of the sum of squares in the DCT domain, and the fact that c_0, the DC DCT coefficient, is eight times the mean of the pixels in the corresponding 8x8 block of pixels. If, for a given macroblock, the intra variance is less than the inter variance, then the macroblock is encoded using intra-encoding techniques (i.e., based on the raw pixel data); otherwise, the macroblock is encoded using inter-encoding techniques (i.e., based on the interframe pixel differences).

Field/Frame Mode Decision in the DCT Domain

Another mode decision supported by the MPEG standard is the field/frame mode decision in which it is determined whether to encode an image as a single frame or separated into the upper and lower fields of a single interleaved image. In conventional MPEG algorithms, the field/frame mode decision is made in the pixel domain. According to embodiments of the present invention, the field/frame mode decision is made in the DCT domain by comparing the variance for the frame DCT mode to the variance for the field DCT mode and selecting the mode having the smaller variance. For DV and MJPEG input

- 17 -

bitstreams, the dequantized DCT coefficient data generated by partial video decoder **102** will already be structured in frame format. As such, the variance Vframe for the frame DCT mode can be computed directly from that dequantized DCT coefficient data using Equations (26) and (27).

According to embodiments of the present invention, the variance for the field DCT mode can be computed using equations similar to Equations (26) and (27), where each 16x16 macroblock X of DCT coefficient data in frame format has been converted to the corresponding 16x16 macroblock Y of DCT coefficient data in field format using the matrix multiplication transformation T of Equation (28) as follows:

$$Y = TX \quad , \tag{28}$$

where the 16x16 transformation matrix T is given by Equation (29) as follows:

$$T = A^t DA, \tag{29}$$

where the 16x16 matrix A is given by Equation (5) and the 16x16 frame-to-field rearrangement matrix D is given by Equation (30) as follows:

$$D = \begin{bmatrix} 1 & 0 \\ & 1 & 0 \\ & & 1 & 0 \\ & & & 1 & 0 \\ & & & & 1 & 0 & & & & \mathbf{0} \\ & & \mathbf{0} & & & 1 & 0 \\ & & & & & & 1 & 0 \\ & & & & & & & 1 & 0 \\ 0 & 1 \\ & 0 & 1 & & & & \mathbf{0} \\ & & 0 & 1 \\ & & & 0 & 1 \\ & & & & 0 & 1 \\ & \mathbf{0} & & & & & 0 & 1 \\ & & & & & & & 0 & 1 \\ & & & & & & & & 0 & 1 \end{bmatrix} \tag{30}$$

- 18 -

Summary of Transcoding Processing

Fig. 4 shows a flow diagram of the DV-to-MPEG transcoding processing performed by transcoding system 100 of Fig. 1, according to one embodiment of the present invention. Transcoding processing begins with partial video decoder 102 of Fig. 1 partially decoding an input encoded video bitstream conforming to the DV standard to generate a set of dequantized DCT coefficients for each frame of DV-encoded data (step 402).

If the input bitstream was encoded in DV 2-4x8 mode, then DCT data converter 104 of Fig. 1 converts the dequantized DCT coefficient data from 2-4x8 mode format into the 8x8 mode format supported by the MPEG standard, according to Equation (18) (step 404). DCT data converter 104 also converts the DCT coefficient data from the YUV 4:1:1 color format into the YUV 4:2:0 color format supported by the MPEG standard, according to Equation (15) (step 406). In addition, DCT data converter 104 optionally resizes the DCT coefficient data by subsampling and applying the anti-aliasing filter, according to Equation (3) (step 408).

Following the processing of DCT data converter 104, partial video encoder 106 of Fig. 1 optionally performs the MPEG field/frame mode decision in the DCT domain, according to Equations (27) and (28) (step 410). Partial video encoder 106 also optionally performs motion estimation at block boundaries in the DCT domain (step 412) in order to support the MPEG inter/intra mode decision also implemented in the DCT domain, according to Equations (26) and (27) (step 414). Partial video encoder 106 requantizes the DCT coefficient data using an optional rate control algorithm, according to Equation (23) (step 416). Partial video encoder then performs conventional MPEG encoding techniques on the requantized DCT coefficient data in order to generate the output encoded video bitstream conforming to the MPEG standard (step 418).

When the input encoded video bitstream conforms to the MJPEG standard, the processing of transcoding system 100 of Fig. 1 is analogous to the processing in Fig. 4, except that steps 404 and 406 are never needed.

The present invention has been described in the context of MJPEG-to-MPEG and DV-to-MPEG transcoding. Those skilled in the art will appreciate that some of the features of the present invention can be applied to transcoding between bitstreams conforming to

different DCT-based image compression algorithms other than only MJPEG or DV as the input bitstream and/or only MPEG as the output bitstream.

The present invention can be embodied in the form of methods and apparatuses for practicing those methods. The present invention can also be embodied in the form of 5 program code embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other machine-readable storage medium, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. The present invention can also be embodied in the form of program code, for example, whether stored in a storage medium, loaded into and/or 10 executed by a machine, or transmitted over some transmission medium, such as over electrical wiring or cabling, through fiber optics, or via electromagnetic radiation, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. When implemented on a general-purpose processor, the program code segments combine with the processor to 15 provide a unique device that operates analogously to specific logic circuits.

It will be further understood that various changes in the details, materials, and arrangements of the parts which have been described and illustrated in order to explain the nature of this invention may be made by those skilled in the art without departing from the principle and scope of the invention as expressed in the following claims.

- 20 -

CLAIMS

What is claimed is:

1   1. A method for converting an input encoded video bitstream conforming to
2   a first DCT-based compression algorithm into an output encoded video bitstream
3   conforming to a second DCT-based compression algorithm different from the first
4   compression algorithm, comprising the steps of:

5   (a) applying decoding steps conforming to the first compression algorithm to the input
6   bitstream to generate dequantized DCT coefficient data in a DCT domain;

7   (b) subsampling and anti-aliasing filtering the dequantized DCT coefficient data in the
8   DCT domain in order to resize the dequantized DCT coefficient data; and

9   (c) applying encoding steps conforming to the second compression algorithm to the
10 resized DCT coefficient data to generate the output bitstream.

1   2. The invention of claim 1, wherein the first compression algorithm is an
2   MJPEG algorithm and the second compression algorithm is an MPEG algorithm.

1   3. The invention of claim 1, wherein the first compression algorithm is a DV
2   algorithm and the second compression algorithm is an MPEG algorithm.

1   4. The invention of claim 1, wherein step (b) comprises the step of applying
2   a subsampling and anti-aliasing filter matrix operation to the dequantized DCT coefficient
3   data in the DCT domain.

1   5. The invention of claim 4, wherein the filter matrix comprises a filter
2   kernel that appears at a different location within each row of the filter matrix.

1   6. An apparatus for converting an input encoded video bitstream conforming
2   to a first DCT-based compression algorithm into an output encoded video bitstream
3   conforming to a second DCT-based compression algorithm different from the first
4   compression algorithm, comprising:

5   (a) means for applying decoding steps conforming to the first compression algorithm to
6   the input bitstream to generate dequantized DCT coefficient data in a DCT domain;

- 21 -

7   (b) means for subsampling and anti-aliasing filtering the dequantized DCT coefficient

8   data in the DCT domain in order to resize the dequantized DCT coefficient data; and

9   (c) means for applying encoding steps conforming to the second compression algorithm

10   to the resized DCT coefficient data to generate the output bitstream.

1   7. A computer-readable medium having stored thereon a plurality of

2   instructions, the plurality of instructions including instructions which, when executed by a

3   processor, cause the processor to implement a method for converting an input encoded

4   video bitstream conforming to a first DCT-based compression algorithm into an output

5   encoded video bitstream conforming to a second DCT-based compression algorithm

6   different from the first compression algorithm, the method comprising the steps of:

7   (a) applying decoding steps conforming to the first compression algorithm to the input

8   bitstream to generate dequantized DCT coefficient data in a DCT domain;

9   (b) subsampling and anti-aliasing filtering the dequantized DCT coefficient data in the

10   DCT domain in order to resize the dequantized DCT coefficient data; and
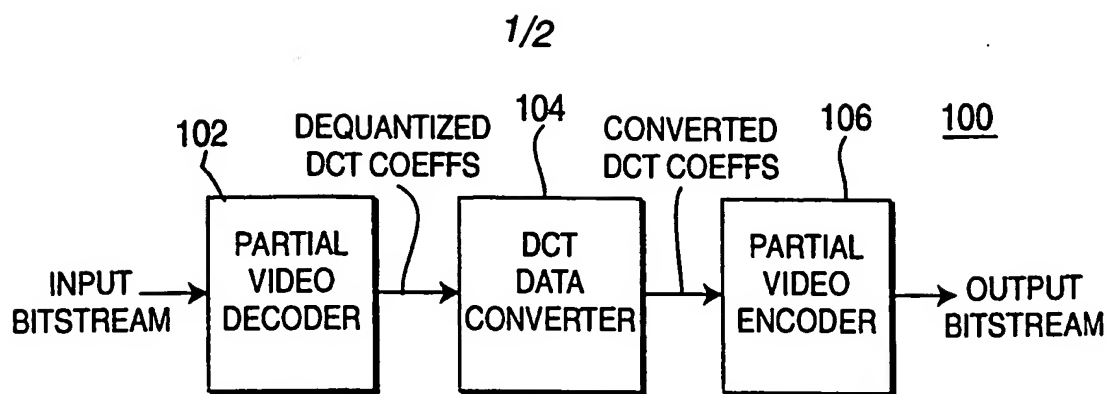
11   (c) applying encoding steps conforming to the second compression algorithm to the

12   resized DCT coefficient data to generate the output bitstream.

1   8. A transcoding system for converting an input encoded video bitstream

2   conforming to a first DCT-based compression algorithm into an output encoded video

3   bitstream conforming to a second DCT-based compression algorithm different from the first

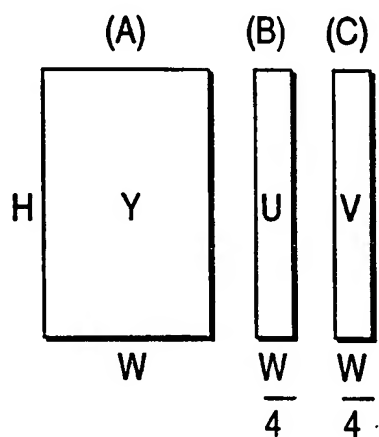4   compression algorithm, comprising:

5   (a) a partial video decoder configured to apply decoding steps conforming to the first

6   compression algorithm to the input bitstream to generate dequantized DCT coefficient data

7   in a DCT domain;

8   (b) a DCT data converter configured to subsample and anti-aliasing filter the

9   dequantized DCT coefficient data in the DCT domain in order to resize the dequantized
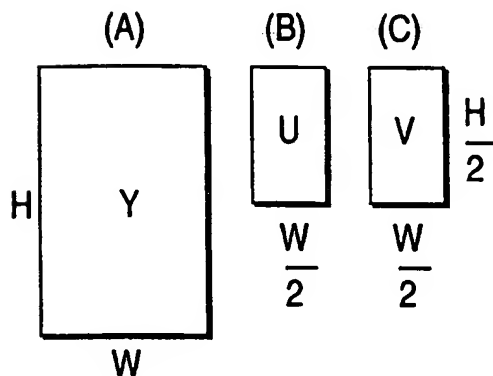
10   DCT coefficient data; and

11   (c) a partial video encoder configured to apply encoding steps conforming to the second

12   compression algorithm to the resized DCT coefficient data to generate the output bitstream.
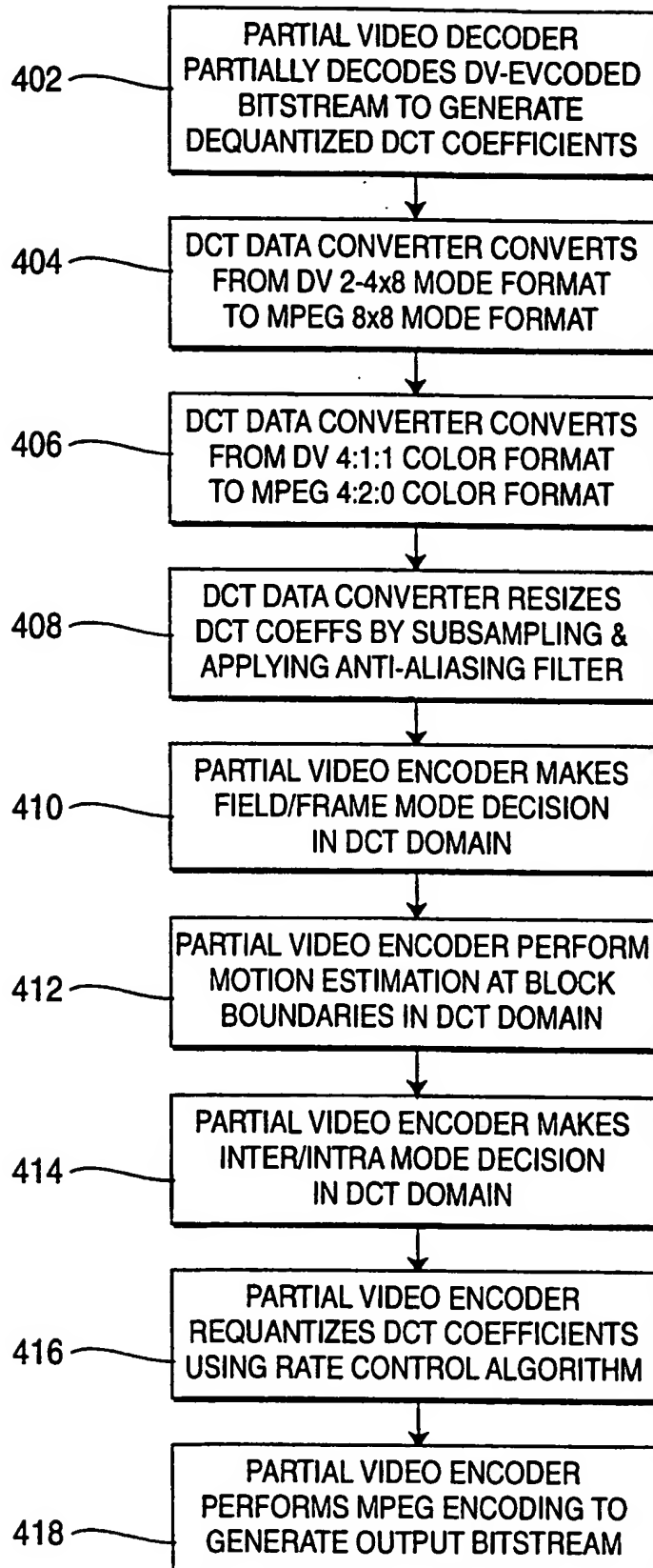
FIG. 1

FIG. 2

FIG. 3

2/2

```
402 ──── PARTIAL VIDEO DECODER
          PARTIALLY DECODES DV-EVCODED
          BITSTREAM TO GENERATE
          DEQUANTIZED DCT COEFFICIENTS
```

```
404 ──── DCT DATA CONVERTER CONVERTS
          FROM DV 2-4x8 MODE FORMAT
          TO MPEG 8x8 MODE FORMAT
```

```
406 ──── DCT DATA CONVERTER CONVERTS
          FROM DV 4:1:1 COLOR FORMAT
          TO MPEG 4:2:0 COLOR FORMAT
```

```
408 ──── DCT DATA CONVERTER RESIZES
          DCT COEFFS BY SUBSAMPLING &
          APPLYING ANTI-ALIASING FILTER
```

```
410 ──── PARTIAL VIDEO ENCODER MAKES
          FIELD/FRAME MODE DECISION
          IN DCT DOMAIN
```

```
412 ──── PARTIAL VIDEO ENCODER PERFORM
          MOTION ESTIMATION AT BLOCK
          BOUNDARIES IN DCT DOMAIN
```

```
414 ──── PARTIAL VIDEO ENCODER MAKES
          INTER/INTRA MODE DECISION
          IN DCT DOMAIN
```

```
416 ──── PARTIAL VIDEO ENCODER
          REQUANTIZES DCT COEFFICIENTS
          USING RATE CONTROL ALGORITHM
```

```
418 ──── PARTIAL VIDEO ENCODER
          PERFORMS MPEG ENCODING TO
          GENERATE OUTPUT BITSTREAM
```

FIG. 4

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   H04N7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KIM D ET AL: "TRANSCODING DV INTO MPEG-2 IN THE DCT DOMAIN" PROCEEDINGS OF THE SPIE,1998, XP000870210 abstract paragraph '02.1!	1-8
A	WU J -L ET AL: "AN EFFICIENT JPEG TO MPEG-1 TRANSCODING ALGORITHM" IEEE TRANSACTIONS ON CONSUMER ELECTRONICS,US,IEEE INC. NEW YORK, vol. 42, no. 3, 1 August 1996 (1996-08-01), pages 447-457, XP000638525 ISSN: 0098-3063 abstract paragraph '02.1! figure 2B	1-8

-/--

[X] Further documents are listed in the continuation of box C.		[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 June 2000	16/06/2000

Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL – 2260 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Authorized officer Georgiou, G

Form PCT/ISA/210 (second sheet) (July 1992)

2

INTERNATIONAL SEARCH REPORT

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	MERHAV N ET AL: "FAST ALGORITHMS FOR DCT-DOMAIN IMAGE DOWN-SAMPLING AND FOR INVERSE MOTION COMPENSATION" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY,US,IEEE INC. NEW YORK, vol. 7, no. 3, 1 June 1997 (1997-06-01), pages 468-476, XP000690585 ISSN: 1051-8215 abstract paragraph '000I! - paragraph '00II!; figure 1B	1-9
A	SHIH-FU CHANG ET AL: "MANIPULATION AND COMPOSITING OF MC-DCT COMPRESSED VIDEO" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS,US,IEEE INC. NEW YORK, vol. 13, no. 1, 1 January 1995 (1995-01-01), pages 1-11, XP000492740 ISSN: 0733-8716 paragraph '000I! paragraphs 'III.D!, paragraph 'III.E!	1-8
A	EP 0 781 052 A (CIRRUS LOGIC INC) 25 June 1997 (1997-06-25) abstract; claims 1-3; figure 3	1-8
A	WO 97 14252 A (SARNOFF DAVID RES CENTER) 17 April 1997 (1997-04-17) abstract; claims 1-3	1-8

2

International Application No
PCT/US 00/04824

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 0781052	A	25-06-1997	US	5832120 A	03-11-1998
			JP	9214972 A	15-08-1997
WO 9714252	A	17-04-1997	EP	0855119 A	29-07-1998
			JP	11511632 T	05-10-1999
			US	5845015 A	01-12-1998